AD-A049 688 PRINCETON UNIV N J DEPT OF STATISTICS F/G 12/1
ON THE EVALUATION OF BIOLOGICAL HAZARDS AND COMPETING RISKS.(U)
SEP 77 G S WATSON N00014-75-C-0453
UNCLASSIFIED TR-132-SER-2 NL

| OF |
AD
A049688

END
DATE
FILMED
3-78
DDC

2

see back page for 1473

ON THE EVALUATION OF BIOLOGICAL HAZARDS AND COMPETING RISKS*

By G. S. Watson
Princeton University

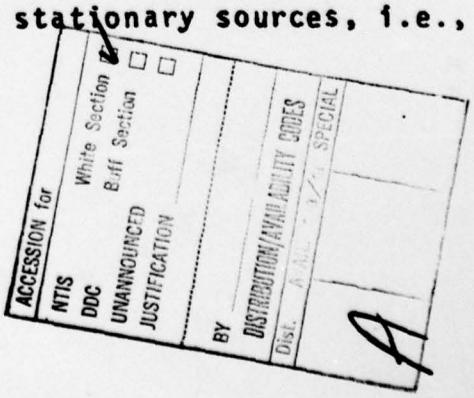
Technical Report No. 132, Series 2
Department of Statistics
Princeton University
September 1977

Research sponsored in part by a contract with the Office of Naval Research, Contract No. N00014-75-C-0453, awarded to the Department of Statistics, Princeton University.

*This report is an expanded version of some extemporaneous remarks at the end of a meeting, with the same title, held at the Institute for Energy Analysis, Oak Ridge Associated Universities, Oak Ridge, Tennessee on September 7-8, 1977.

Professors Birnbaum and Neyman have elected me, (as they assert my native language is English, an honor rarely accorded to Australians), to thank the organizers and to suggest that this dialogue between biologists and statisticians be continued in future meetings. It is painfully clear that we have only scratched the surface of the problem. This problem is: how to understand environmental hazards to health through experimental and epidemiological research, and how to provide wise advice to those who must frame policies for regulation. It is truly an enormous task.

Oak Ridge is a natural place for such a meeting since much of the environmental problem arises from the demand for more energy and, of course, Oak Ridge owes its existence to nuclear energy. The most studied hazard is radiation and the largest animal experiments have been conducted here. But chemicals are an increasing menace to health. In particular, the use of fossil fuels to generate energy raises many medical problems. Thus the Institute for Energy Analysis at Oak Ridge is a natural host for this meeting. Of course, such work is being done by many groups, e.g., the National Research Council has a committee considering the "Research Needs for the Health Effect of Fossil Fuels." Its charge is restricted to the effects of stationary sources, i.e., power plants.



1. INTRODUCTION - COMMENTS ON THIS CONFERENCE

I have been asked to make my observations on this meeting. The papers have reviewed some literature, discussed some experiments and their analysis, some ideas have been suggested that are worth following up, and many essential things have not been mentioned.

Speakers have continually referred to cancer. This is not the greatest public health problem! I believe that the end of life is less important than the quality of life, which means that I think chronic diseases should have been the main topic. When we leave radiation and turn to chemicals (e.g., the effects of effluents from a fossil fuel power station), we think of debilitating things such as asthma, etc. And yet everyone has been concerned with the causes of death and the length of life.

Below I have given my views on a number of issues. My major points are -

(i) there are no competing risks in the ILLNESS - DEATH MODEL

(ii) it seems very hard to frame a biologically reasonable model to make sense of the potential lifetime concept, but an attempt is made in §6.

2. CAUSES OF DEATH

For human populations the causes of death usually are very hard to determine. We have seen that even with the most lavish studies of mice, they are equally vague. Thus Mr. Neyman has very sensibly tried to persuade us that we should try to work with what is a little better determined---the medical description of the state of the deceased mouse. For humans who have not been under con-

tinuous good medical care, even this will be vague, too. With computers, such complex information can be studied. More empirical data-analytical methods (see, e.g., the new book by Gnanadesikan) will have to be used.

The statisticians present learned about the great contrasts between animals (usually mice) and humans---the permissible conduct of experiments and the causes of death. The extrapolation from animal experiments to man is agreed to need a real understanding of causes and mechanisms.

3. COMPETING RISKS

Much of our time was spent on "The Competing Risks" problem in different ways and at different mathematical levels and always with different notations! The basic problem seems clear with no mathematics at all. Let us assume that there is a list of C (primary) causes of death. Classify and count all deaths in a large human population in a year and divide each number by the population size. If the first cause of death is eliminated, it seems <u>obvious</u> that, with no further assumptions, there is no way of predicting the C-1 ratios that will be seen in the future. We simply don't know what will be the fate of those who formerly died from the first cause.

Biologically, we know that the <u>susceptibility</u> of an individual to various pathological states (as well as eye colors, etc.) will depend upon his/her genotype. The <u>incidence</u> of these conditions will depend upon the life history of the individual, i.e.,

the individual's environment in the largest sense of the word. In a population there is variation in both genotype and environment. At least in principle these two sources of variation would have to determine the joint distribution of the potential times to death from the various causes of an individual chosen at random from this population, if we can make sensible definitions of these potential lifetimes (see §6). Now, the genetic component is weaker in some conditions than in others. If also there is little correlation between individual environments, we may expect "potential life times" for these conditions to be roughly independent. If the genetic background is homogeneous (e.g., an inbred mice strain), we need only worry about the environmental correlations. Conversely, if the environment is constant, only the genetic variation need be considered.

Everyone knows of the difficulties in the Nature-Nurture argument---I.Q. and race is a notorious example. Even identical twin studies may be criticized. The genetics of quantitative characters is simply not in a state to help us.

The several paragraphs above were written in the hope that they will throw some biological light on the statistician's discussion of "independence" and "dependence" in relation to "Competing Risks." Recall also that the notion of a cause of death is easily criticized. Thus the statisticians' formulation of the problems, however amusing it may seem to them, may not be very good. That it leads to conundrums like "non-identifiability" may not be a cause for tears. The problem is real enough however. And yet I

have never been able to convince medical people that it might upset their conclusions that a treatment was reducing a death rate, for example! I have the feeling that this problem may be over-emphasized. The theory appears also under the subject heading of Reliability Theory. However, I am convinced this latter theory has not made any technical devices more reliable, though many of us have enjoyed writing and reading papers about it.

I will return to this argument in §6.

4. ON HERETICS AND HEROES

I am personally attracted to people who rationally advocate non-establishment views. Two come to mind in this general area and I mention them in case they are unfamiliar to some. P. R. J. Burch has written two books and many papers in which the ages of incidence or onset of many conditions has been studied. He seems to have done more work in this area than anyone else. In a recent letter, he charmingly described himself as a "recognized heretic." While well aware of the many difficulties with medical data, he is not unduly worried about his curves being upset by competing risks.

I turn now to cancer to make several points. In this area, Armin Braun is something of a heretic. His book, The Biology of Cancer, makes very good reading. He does not accept the almost universal equating of mutation and cancer. (See also my note in P.N.A.S., April 1977). Certainly all speakers at this meeting made this assumption---of course it may well be true in these specific

cases--osteosarcoma due to bone-incorporated strontium 90 and melanoma due to U.V. in sunlight. Other cases where it is an attractive hypothesis are the early childhood cancers like retinoblastoma. The models presented have a natural appeal to an applied mathematician/statistician. But they seem to raise as many questions as they answer. (Of course, this is characteristic of any good theory.) One point that worries me is the following.

In the resting state the DNA in a nucleus is a supercoiled, tangled mass like a full pot of spaghetti. To take the Groer-Marshall theory, an α particle going through this mess is likely to do a lot of damage. Some of this damage is supposed to lead to a mutation which is a step to carcinogenesis, some is to be repaired and some to kill the cell. It seems intuitively to me that some damage will be irreparable, and some will be repaired, often with error so that many mutations will be produced. The progeny of some mutated cells will not survive the competition with normal and other mutated cells. Many mutations may be synonomous or selectively neutral and so for growth purposes be equivalent to normal cells. Finally some mutated cells are freed from growth restraints and become cancer cells. Has anyone looked for mutants among surviving cells and sought to relate their prevalence to dosage? For they should obey the same laws as cancer-transformed cells. Such mutants should be detectable electrophoretically.

There is a large literature no doubt examining critically the relative sizes of DNA molecules and their radiation cross-

sections. It is a little dangerous to think about cloud chamber tracks which are visible to the naked eye and make deductions about molecules.

Since it is useful to read into the record of conferences references to highly relevant work, I would like to mention the P.N.A.S. 1976 paper by Neyman and Puri on a model for radiation effects and a 1974 book Intrinsic Mutagenesis by Sir Macfarlane Burnet which is concerned with the whole problem of this conference.

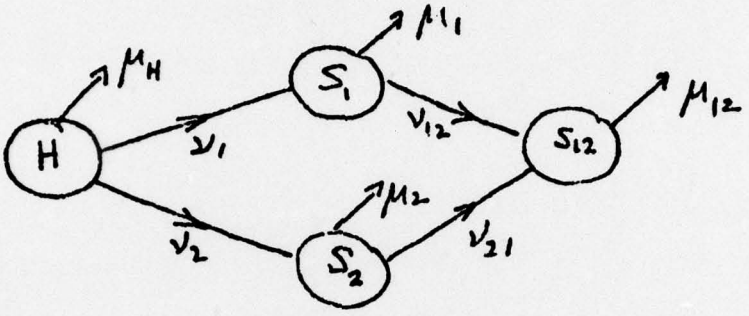
5. ILLNESS AND DEATH MODELS

These models stemming ultimately again from Mr. Neyman, have been discussed by C.L. Chiang here and in his well-known book. In passing the recent result of P. Clifford (P.N.A.S., 1977) was alluded to. This is such an important comment on such models that it is a pity it was not discussed at length. Hence I will give a brief summary.

The random progression of the health state of an organism might well be a Markov process if the description of the state is complex enough. (There was some discussion of this and it seems that in practice the state designation is too simple for this to be valid.) There will be many transition functions to be estimated in any actual experiment. In assessing the effect of a changing environment we will be particularly interested in transition functions that are time dependent.

If one considers the simplest conceivable such model, Clifford

shows that the largest of survival and sacrifice experiments will lead to ambiguity! Let H stand for healthy, S_i for having disease i,(i=1,2) and S_{12} for having both diseases. Suppose only forward transitions are possible and that death can occur in any of the four live states. The following figure should now be clear.



If $P_h(t)$, $P_{s_1}(t)$ etc. are the probabilities of being in live states H, D, etc., at time t, then with a dot for d/dt.

$$
\begin{aligned}
\dot{P}_h &= -\mu_h P_h - \nu_1 P_h - \nu_2 P_h \\
\dot{P}_{s_1} &= -\mu_1 P_{s_1} + \nu_1 P_h - \nu_{12} P_{s_1} \\
\dot{P}_{s_2} &= -\mu_2 P_{s_2} + \nu_2 P_h - \nu_{21} P_{s_2} \\
\dot{P}_{s_{12}} &= -\mu_{12} P_{s_{12}} + \nu_{12} P_{s_1} + \nu_{21} P_{s_2}
\end{aligned}
\qquad (1)
$$

Denoting the death states by D_h, D_1, D_2, D_{12}, we have similarly

$$\dot{P}_{d_h}(t) = \mu_h P_h \ , \ \dot{P}_{d_1} = \mu_1 P_{d_1}$$
$$\dot{P}_{d_2} = \mu_2 P_{d_2} \ , \ \dot{P}_{d_{12}} = \mu_{12} P_{d_{12}} \qquad (2)$$

Now with an enormous (i.e., "gedanken") experiment and observing and sacrificing at very fine time intervals, we can determine all the P's as functions of time so that equations (2) determine the μ's. Writing $\dot{P}_h + \mu_h P_h = R_h$, etc., which <u>are now known functions of time</u>, (1) becomes

$$\begin{bmatrix} R_h \\ R_{s_1} \\ R_{s_2} \\ R_{s_{12}} \end{bmatrix} = \begin{bmatrix} -P_h & -P_h & 0 & 0 \\ P_h & 0 & -P_{s_1} & 0 \\ 0 & P_h & 0 & -P_{s_2} \\ 0 & 0 & P_{s_1} & P_{s_2} \end{bmatrix} \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_{12} \\ \nu_{21} \end{bmatrix} . \qquad (3)$$

These equations may be written $R = PV$ where P clearly has a zero determinant. Thus there is no unique solution for the μ's, although $\nu_1 + \nu_2$ is determined.

The "potential life time" model asserts that with Causes of death, there is a joint distribution of potential times to death $T_1, \ldots, T_C$ and that the time of death T is given by

$$T = \min(T_1,\ldots,T_C) \ . \qquad (4)$$

This makes no sense whatever in this model. But even more striking, there are no competing risks in this model! For when the individual is in H, a "healthy death" is the only cause, when in S_1, this is the only cause of death, etc. Thus if we follow Mr. Neyman's suggestion of giving a pathological description, and there are no back flows, these models shed no light whatever on competing risks.

6. COMPETING RISKS AGAIN

I would like to develop the genetical picture in §3 to give yet another account of the competing risk problem. My excuse is that I believe it throws some new light on two aspects of the issue, (i) the definition of "potential life times," (ii) the question of dependence. This work was done after the conference and for most of this period I was convinced that current formulation involving (4) made no sense whatever!

For any individual i, we suppose that the chance that he dies from cause c in $(t,t+dt)$ is $\mu(t,c,i)dt$ where $c = 1,2,\ldots,C$. If only cause C were operating, he would have a lifetime $T_c(i)$, say, where

(1) $$\text{Prob}\,(T_c(i)\epsilon(t,t+dt)) = \exp\left\{-\int_0^t \mu(t',c,i)dt'\right\}\mu(t,c,i)dt$$

and

(2) $$\text{Prob}(T_c(i) > t) = \exp\left\{-\int_0^t \mu(t',c,i)dt'\right\}.$$

We call $T_c(i)$ his potential lifetime for cause of death c. They are clearly non-observable since several causes act simultaneously.

When all C causes are competing for his life, define

(3) $$\mu(t,i) = \sum_{c=1}^{C} \mu(t,c,i) \ . \quad \text{Then}$$

$$P(\text{ i dies of cause } c \text{ at an age in } (t, t+dt))$$

$$(4) \qquad = \exp\left(-\int_0^t \mu(t',i)dt'\right)\mu(t,c,i)dt \ ,$$

$$(5) \qquad = \prod_{\substack{c'=1 \\ c' \neq c}}^{C} \text{Prob}(T_{c'}(i) > t+dt)\,\text{Prob}(T_c(i)\varepsilon(t,t+dt)) \ ,$$

$$(6) \qquad = \text{Prob}\left(T_c(i) = \min(T_1(i),\ldots,T_C(i)), T_c(i)\varepsilon(t,t+dt)\right).$$

(5) follows from (4) by (1) and (2) on observing that we need only the term of order 1. (5) and (6) make it clear that with this model the individual's potential lifetimes may be regarded as independent and he dies at the age of the least one. Further, the chance of death for i in (t,t+dt) is

$$\sum_{c=1}^{C} \exp\left(-\int_0^t \mu(t',i)dt'\right)\mu(t,c,i)dt = \exp\left(-\int_0^t \mu(t',i)dt'\right)\mu(t,i).$$

Setting $1 - F_i(t) = \text{Prob}(\text{ i living longer than } t)$, we have

$$(7) \qquad 1 - F_i(t) = \exp\left(-\int_0^t \mu(t',i)dt'\right) \ .$$

Now suppose that the population contains a fraction p_i of individuals like i. For example, this group all have the same genotype and environmental history. Then the life distribution in the population is

$$(8) \qquad \begin{cases} F(t) = \Sigma p_i F_i(t) \\ \qquad = 1 - \Sigma p_i \exp\left(-\int_0^t \mu(t',i)dt'\right) \\ \qquad = 1 - \Sigma p_i \prod_{c=1}^{C} P(T_c(i) > t) \end{cases}$$

and probability density of life spans terminated by cause c, L_c, say, at t is the average of (4) or (5) or (6),

$$(9)\qquad P\left(L_c \epsilon (t,t+dt)\right) = \Sigma p_i\ \mu(t,c,i)\exp\left(-\int_0^t \mu(t',i)dt'\right)dt$$

$$= \Sigma p_i\ P\left(T_c(i)\epsilon(t,t+dt)\right) \prod_{c' \neq c} P\left(T_{c'}(i) > t\right)dt$$

while

$$(10)\qquad P(L_c > t) = \int_t^\infty P\left(L_c \epsilon (t'',t''+dt'')\right).$$

Thus one cannot treat the L_c's as independent variables and use their minimum as we could the $T_c(i)$'s. The L_c's could conceivably be called the <u>potential lifetimes of a randomly chosen individual</u>. If the "acting alone aspect" is valued we might consider instead $D_1,\ldots,D_c$ where

$$\text{Prob}\left(D_c \epsilon (t,t+dt)\right) = \sum_i p_i\ \mu(t,c,i)\exp\left(-\int_0^t \mu(t',c,i)dt'\right)$$

$$\text{Prob}(D_c > t) = \sum_i p_i\ \exp\left(-\int_0^t \mu(t',c,i)dt'\right).$$

These do not seem to be good candidates.

If one cannot classify individuals, a detailed study of the population will only reveal

$$(11)\qquad \left\{(dt)^{-1}\cdot \begin{array}{l}\text{fraction of individuals}\\ \text{dying of } c \text{ in } (t,t+dt)\end{array}\right\} = \Sigma p_i\ \mu(t,c,i) = \bar{\mu}(t,c,), \text{say},$$

$$(12)\qquad \left\{(dt)^{-1}\cdot \begin{array}{l}\text{fraction of individuals}\\ \text{dying in } (t,t+dt)\end{array}\right\} = \sum_c \bar{\mu}(t,c), = \bar{\mu}(t), \text{ say}.$$

From (12) we will assert that the life distribution is

$$(13)\qquad \bar{F}(t) = 1 - \exp\left(-\int_0^t \bar{\mu}(t')dt'\right).$$

There is <u>always</u> a function $\bar{\mu}(\cdot)$ such that

$$(14)\qquad \exp\left(-\int_0^t \bar{\mu}(t)dt\right) = \sum_i p_i \exp\left(-\int_0^t \mu(t',i)dt'\right)$$

as may be seen explicitly by taking logarithms and differentiating. Similarly, we will write, instead of (9),

$$\exp\left\{-\int_0^t \bar{\mu}(t',c)dt'\right\}\bar{\mu}(t,c) \ .$$

It is clear that even if the p_i are known (but not which individuals belong to type i), knowledge of $\bar{\mu}(t,c)$ does not determine the $\mu(t,c,i)$.

With this set up, suppose cause C is eliminated. This means that $\mu(t,C,i) \equiv 0$, for all i,t. Equations (1) to (7) change very simply. For example, (7) now reads

$$(15) \quad 1 - F_i^*(t) = \prod_{c=1}^{c-1} P(T_c(i) > t) \ ,$$

so that (8) now reads

$$(16) \quad 1 - F^*(t) = \Sigma p_i \prod_{c=1}^{c-1} P(T_c(i) > t) \ .$$

Thus (16) cannot be determined unless the individuals can be classified.

If they can be classified, all the problems disappear with a sufficiently large experiment.

REFERENCES

Braun, A. (1974). The Biology of Cancer. Reading, Mass.: Addison-Wesley.

Burch, P. R. J. (1968). Growth, Disease and Aging. Edinburgh, Scot.: Oliver & Boyd.

_____ (1976). The Biology of Cancer: A New Approach. Lancaster, U.K.: M.T.P. Press.

Burnet, Sir Macfarlane, (1974). Intrinsic Mutagenesis. New York: John Wiley & Sons.

Clifford, Peter. (1977). Non-identifiability in stochastic models of illness and death, Proc. Nat. Acad. Sci., 74, 1338-40.

Gnanadesikan, R. (1977). Methods for Statistical Data Analysis of Multivariate Observations. New York: John Wiley & Sons

Neyman, J. & Puri, P. S. (1976). A structural model of radiation effects in living cells, Proc. Nat. Acad. Sci., 73, 3360-63.

Watson, G. S. (1977). Age-incidence curves for cancer, Proc. Nat. Acad. Sci., 74, 1341-42.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(14) TR-132-SER-2

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report No. 132 ✓
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): (6) ON THE EVALUATION OF BIOLOGICAL HAZARDS AND COMPETING RISKS.
5. TYPE OF REPORT & PERIOD COVERED: (9) Technical rept.,
6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s): (10) G.S. Watson
8. CONTRACT OR GRANT NUMBER(s): (15) N000 14-75-C-0453 ✓

9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics ✓ Princeton University, Princeton, N.J.
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research (Code 436) Arlington, Va. 22217
12. REPORT DATE: Sep[redacted]77 (11) (12) 16p.
13. NUMBER OF PAGES: 14

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES: This report is an expanded version of some extemporaneous remarks at the end of a meeting (with the same title), held at the Institute for Energy Analysis, Oak Ridge Assoc. Univ. Oak Ridge, Tennessee, on September 7-8, 1977.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Competing risks, illness/death models, carcinogenesis

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

It is difficult to find a biological basis for the competing risk model and potential lifetimes but an effort is made ~~here~~ in this paper. Clifford's results on illness and death models are described. Miscellaneous remarks are made on several other related topics.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED 406 873
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)